Crystal Structure and Directed Evolution of Specificity of NlaIV Restriction Endonuclease

Honorata Czapinska[1], Wojciech Siwek[1,4], Roman H. Szczepanowski[1], Janusz M. Bujnicki[1,2], Matthias Bochtler[1,3] and Krzysztof J. Skowronek[1]

*1 - **International Institute of Molecular and Cell Biology,** Trojdena 4, 02-109 Warsaw, Poland*
*2 - **Institute of Molecular Biology and Biotechnology,** Faculty of Biology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznan, Poland*
*3 - **Institute of Biochemistry and Biophysics PAS,** Pawinskiego 5a, 02-106 Warsaw, Poland*

***Correspondence to Janusz M. Bujnicki, Matthias Bochtler and Krzysztof J. Skowronek:** International Institute of Molecular and Cell Biology, Trojdena 4, 02-109 Warsaw, Poland. iamb@genesilico.pl, mbochtler@iimcb.gov.pl, kskowronek@iimcb.gov.pl*
https://doi.org/10.1016/j.jmb.2019.04.010
Edited by Dan Tawfik

Abstract

Specificity engineering is challenging and particularly difficult for enzymes that have the catalytic machinery and specificity determinants in close proximity. Restriction endonucleases have been used as a paradigm for protein engineering, but successful cases are rare. Here, we present the results of a directed evolution approach to the engineering of a dimeric, blunt end cutting restriction enzyme NlaIV (GGN/NCC). Based on the remote similarity to EcoRV endonuclease, regions for random mutagenesis and *in vitro* evolution were chosen. The obtained variants cleaved target sites with an up to 100-fold k_{cat}/K_M preference for AT or TA (GGW/WCC) over GC or CG (GGS/SCC) in the central dinucleotide step, compared to the only ~17-fold preference of the wild-type enzyme. To understand the basis of the increased specificity, we determined the crystal structure of NlaIV. Despite the presence of DNA in the crystallization mix, the enzyme crystallized in the free form. We therefore constructed a computational model of the NlaIV–DNA complex. According to the model, the mutagenesis of the regions that were in the proximity of DNA did not lead to the desired specificity change, which was instead conveyed in an indirect manner by substitutions in the more distant regions.

Introduction

Specificity engineering is a challenging task. Type II restriction endonucleases provide a good testing ground for methods to design or evolve specificities. The altered cleavage preferences can not only be easily assayed but also be exploited with selection schemes that use the activity to enrich or deplete DNA templates for enzymes that have desirable properties [1]. Despite the advantages of restriction endonucleases in such tasks, successes of rational or semi-rational specificity engineering are rare [1–5]. This results in part from the close spatial proximity of catalytic sites and specificity determining regions, which makes enzyme activity vulnerable to changes intended to alter specificity, and in part from many interdependent degrees of freedom of the substrate DNA, which make detailed modeling difficult.

The restriction endonuclease NlaIV from *Neisseria lactamica* is a dimeric enzyme that belongs to the PD-(D/E)XK class of endonucleases. The only structural information on NlaIV currently available is based on computational modeling [6]. Because of the low amino acid similarity between the enzyme and templates, the model is of relatively low resolution. Nevertheless, some of its key features, such as the predicted active site and some DNA binding residues have already been validated by site-directed mutagenesis [6].

The enzyme cleaves DNA within the GGN/NCC sites to blunt-ended products [7]. The central dinucleotide step of the target sequence provides an opportunity to engineer additional specificity. As

Table 1. Cleavage kinetics of the selected variants

Variant	Sequence	k_{cat} (min^{-1})	AT/GC ratio	K_m (nM)	AT/GC ratio	k_{cat}/K_m (s^{-1} M^{-1})	AT/GC ratio
Wild type	GG**AT**CC	0.331 (±0.038)	2.88	6.4 (±0.5)	0.17	8.69×10^6	17.01
	GG**GC**CC	0.115 (±0.008)		37.5 (±1.3)		5.11×10^5	
I36P/K38R	GG**AT**CC	3.890 (±0.270)	231.55	40.2 (±5.0)	2.38	1.61×10^7	97.58
	GG**GC**CC	0.017 (±0.001)		16.9 (±2.6)		1.65×10^5	
N178A/K180R	GG**AT**CC	0.152 (±0.013)	25.46	22.5 (±2.0)	0.20	1.13×10^6	126.40
	GG**GC**CC	0.006 (±0.000)		111.2 (±15.2)		8.94×10^3	
S176A/K179A/K180W	GG**AT**CC	1.010 (±0.090)	27.60	19.4 (±1.0)	0.69	8.68×10^6	40.19
	GG**GC**CC	0.037 (±0.002)		28.3 (±2.5)		2.16×10^5	

Error values are standard deviations from triplicate experiments.

NlaIV is dimeric, it recognizes only "half" of its target independently and the interactions with the other half are enforced by the 2-fold symmetry of the enzyme-DNA complex [6]. Therefore, the specificity can only be evolved or engineered for one of the two central positions in the target sequence, and the selectivity for other position is symmetry imposed. Hence, only four non-degenerate specializations of the 5′-NN-3′ dinucleotide step are possible, namely AT, TA, GC and CG. If binary ambiguities are allowed, the YR, RY, SS and WW steps become possible, where Y stands for pyrimidine, R for purine, S ("strong") for G or C and W ("weak") for A or T. In principle, there is also the possibility of triply degenerate specificity (e.g., H for A, C or T), but recognition of all but one base in any given position is difficult to implement mechanistically and rarely found in natural restriction enzymes. Hence, this possibility will not be discussed further.

Here, we present an adaptation of the compartmentalization strategy for *in vitro* protein evolution [8–10] to specificity engineering of restriction endonucleases. Using NlaIV as an example, we performed a selection strategy that enhanced the natural preference of the enzyme for an AT over a GC step. Although the positive selection was applied only for cleavage of AT steps, the evolved enzymes also cleave targets with central TA steps. Thus, the engineering created a new endonuclease with specificity for palindromic GGW/WCC target sequences, which is novel and not known to occur for any naturally evolved enzyme. We also present a crystal structure of NlaIV that demonstrates that the changes made to alter its specificity are likely to be positioned away from the target DNA, act indirectly and not by specific interactions with the DNA bases.

Results

Expression and characterization of NlaIV

As a starting point for the *in vitro* evolution of NlaIV specificity, we tested whether the enzyme had an intrinsic preference for the central dinucleotide step. A detailed kinetic characterization revealed that the k_{cat}/K_M value, a measure of the enzyme efficiency, was about 17-fold higher for the GGATCC than for the GGGCCC target sequence (Table 1). The higher k_{cat}/K_M value was a result of approximately 6-fold tighter binding and a 3-fold faster catalytic rate. The finding suggested that it may be possible to build on this preference and evolve NlaIV variants with genuine GGATCC specificity.

Design of the library of NlaIV variants

Sites for mutagenesis were chosen based on the published model of the NlaIV in complex with target DNA [6]. Altogether, five partially overlapping regions (here labeled from A to E) were selected, and in each region, the codons for 2–5 amino acids were altered (Table 2 and Supplementary Fig. S1). For many codons, only the distinction between a pyrimidine and a purine base in the third position is relevant for the coding capacity. Therefore, triplet randomization with NNS instead of NNN reduces the complexity of the library at the nucleic acid level 2-fold, without altering the range of amino acids that can be encoded and decreasing stop codon frequency from 3/64 (0.047) to 1/32 (0.031). Such reduction of the genetic code results also in more even distribution of codons for all amino acids. For instance, the relative frequency of methionine and tryptophan codons increases from 0.016 (1/64) to 0.031 (1/32). To increase the likelihood of substitutions in promising sites at the expense of substitutions in less promising sites, the extent of the

Table 2. Regions selected for mutagenesis

Region	Residues selected for mutagenesis (substitution frequency)
A	I36 (1.0), K38 (0.5), K40 (0.5)
B	K38 (1.0), K40 (0.3), D41 (1.0), G44 (0.3), N45 (0.3)
C	D41 (1.0), G44 (0.5), N45 (0.5)
D	T68 (1.0), Q69 (1.0)
E	S176 (0.8), N178 (0.8), K179 (0.8), K180 (0.8)

Desired substitution frequencies are given in brackets.

mutation rate was separately chosen for each position, according to an estimate of the probability that it will contribute to specificity (Table 2). By a "split and mix" strategy, the codon triplet for the wild-type amino acid was mixed with the NNS codon in a predefined, position-specific ratio. Sequencing of the two of the libraries confirmed that this procedure generated approximately the expected ratios of original and altered codons (Supplementary Table S1).

Selection system for the NlaIV activity

For *in vitro* evolution of NlaIV toward variants of altered specificity, we adopted the strategy of directed evolution by compartmentalization [8–10]. The DNA for wild-type or mutated NlaIV was embedded in droplets of a water in oil emulsion system, ideally in such a manner that there was only a single molecule of DNA with the coding frame in each droplet. The template was transcribed and translated in the droplet, and acted on itself. We embedded the NlaIV coding sequence with T7 promoter in between two restriction sites, a GGSSCC and a GGATCC site. The GGSSCC site was preceded by an upstream primer sequence, and the GGATCC was followed by a downstream primer binding site. The biotinylation at the 5′ end of the non-coding strand was introduced using a biotinylated reverse primer in the PCR run. After transcription and translation in droplets, the emulsion was broken and the DNA attached to biotin was removed by pull-down using streptavidin beads. This step ensured that only the open reading frames coding for NlaIV variants able to cleave the GGATCC sequence, and sever the open reading frame from the biotin tag, were retained (Fig. 1 and Supplementary Fig. S2).

Subsequently, a two-step PCR reaction of the template mixture was carried out, that reintroduced the complete cleavage sites and the biotin tag. The upstream PCR primer for the first reaction was designed in such a way that only the NlaIV open reading frames that were not severed by cleavage of the GGSSCC site could be amplified. Protein expression in the droplets, biotin pull-down and PCR were reiterated five times. Note that in any given drop, only one of the possible SS sequences, that is, GG, GC, CG or CC, is present. However, when the procedure is reiterated, a new implementation of the SS consensus is chosen in each round. Therefore, eventually all four variants of the SS sequence were potentially selected against. The selection was carried out independently for the five A-E pools, each representing a set of NlaIV variants with mutagenesis sites concentrated in a particular region of the protein (Fig. 1 and Supplementary Fig. S2).

The selected NlaIV variants were initially screened against phage λ DNA. Variants that generated cleavage patterns distinct from the wild-type enzyme were further characterized on a panel of short DNA substrates containing a single GGNNCC recognition site. Without the requirement of palindromic symmetry, there are 16 possibilities for the central 5′-NN-3′ step. Four of them, AT, TA, GC and CG, are compatible with 2-fold symmetry. Each of the other 12 dinucleotide steps has a non-identical reverse complement. Therefore, of the remaining 12 steps, only 6 have to be tested, which brings the number of targets to a total of 4 + 6 = 10 (Fig. 2).

Among the five pools of NlaIV variants, pools A (potential alterations of I36, K38, K40), D (T68, Q69) and E (S176, N178, K179, K180) proved successful. From the transformation of DNA representing these pools, single colonies that exhibited altered NlaIV activity could be isolated. These were I36P/K38R, I36R/K40R, W68T/V69Q, S176A/K179A/K180W, N178A/K180R and S176R/N178V/K179A/K180A. We also constructed a variant with two altered regions combined (I36P/K38R/N178A/K180R). Of note, at least two mutations were required to alter specificity. Interestingly, all variants not only showed specificity against the central SS steps and for the AT step, as expected based on the selection strategy, but also continued to cleave targets with the TA step at the center. This could indicate either that the positive selection for the GGATCC specificity was not very effective, which would lead to GGWWCC preference by anti-selection against GGSSCC, or that the mechanistic constraints of the AT step recognition could automatically favor the TA step as well.

Although NlaIV is dimeric, one variant of the enzyme from the E-pool, N178A/K180R, also cleaved DNA with the non-palindromic target sequence GGTGCC. Cleavage of "slightly" non-palindromic DNA by a dimeric restriction endonuclease is not hard to explain

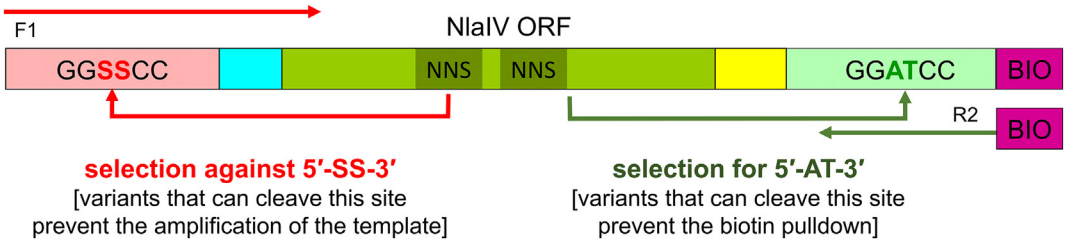

Fig. 1. Selection strategy of NlaIV variants with altered specificity.

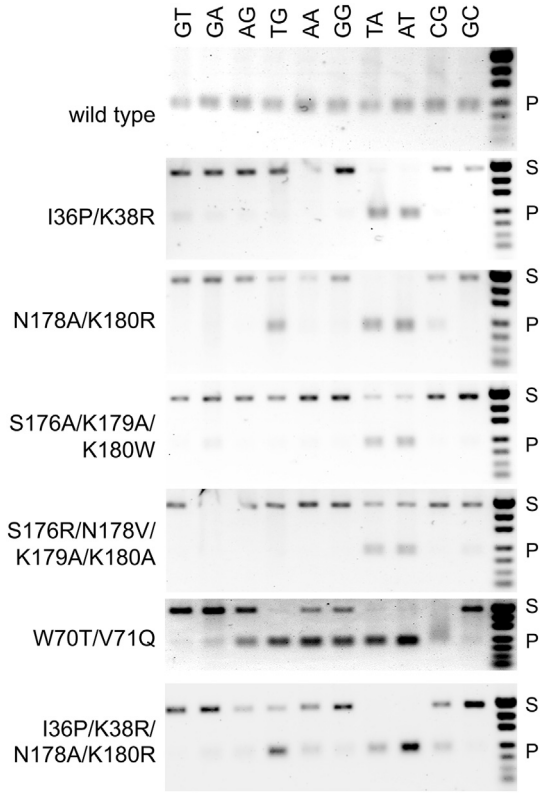

Fig. 2. Initial screening of the selected NlaIV variants. NlaIV variants were screened on a panel of ten 444-bp-long substrates with all dinucleotide combinations in the center of NlaIV target site.The substrate (0.1 μg) was digested with 0.5 μl of enzyme variants (equivalent of 180 μl of the induced culture, on average 50 ng of the enzyme) for 1 h. DNA size standard, pUC19 DNA digested with MspI; S, substrate; P, product.

for one strand. However, since an asymmetric binding mode has to result from the violation of the 2-fold symmetry in the target, it is surprising that both DNA strands are cleaved. For now, we lack a clear explanation for the mechanistic basis of this observation. Double-strand cleavage of symmetry breaking, asymmetric oligoduplexes by dimeric restriction endonucleases otherwise imposing 2-fold symmetry has not been reported to our knowledge for naturally occurring enzymes. However, it is possible that such activity has been overlooked because of an expected specificity for palindromic targets.

Kinetic studies of selected NlaIV variants

Four NlaIV variants exhibited the clearest change of substrate preference with respect to the wild-type enzyme (I36P/K38R, S176A/K179A/K180W, N178A/K180R, S176R/N178V/K179A/K180A). They were therefore further characterized by measuring cleavage rates of two fluorescent substrates with GGATCC and GGGCCC NlaIV target sites at 50 nM concentrations (Supplementary Figs. S3 and S4). Interestingly, the combination of changes in two regions in one NlaIV variant (I36P/K38R/N178A/K180R) did not cause further improvement of the enzyme specificity (Fig. 2 and Supplementary Fig. S4). We determined k_{cat} and K_M values for three of the variants with the most promising properties (Table 1). The kinetic parameters indicated that engineering had a moderately detrimental effect on the binding of DNA with the GGATCC sequence (Fig. 3). The K_M value for DNA with the anti-selected GGGCCC sequence, already lower in the wild-type enzyme, deteriorated only for one of the variants (N178A/K180R), whereas it improved for the other two. The enhanced preference of the variants for GGATCC over GGGCCC was therefore predominantly a k_{cat} effect. Altogether, the 17-fold preference of the wild-type NlaIV for GGATCC over GGGCCC was moderately improved to between ~40- and ~130-fold. This increase in specificity did not always come at the cost of a reduced k_{cat}/K_M value for the desired GGATCC target. In at least one instance of the I36P/K38R variant, the k_{cat}/K_M was improved nearly 2-fold, whereas it stayed similar to the value for the wild-type enzyme for the S176A/K179A/K180W variant and was 10-fold reduced for the N178A/K180R variant.

Crystal structure of NlaIV

To understand the structural basis of altered specificity of NlaIV variants, we aimed to solve the

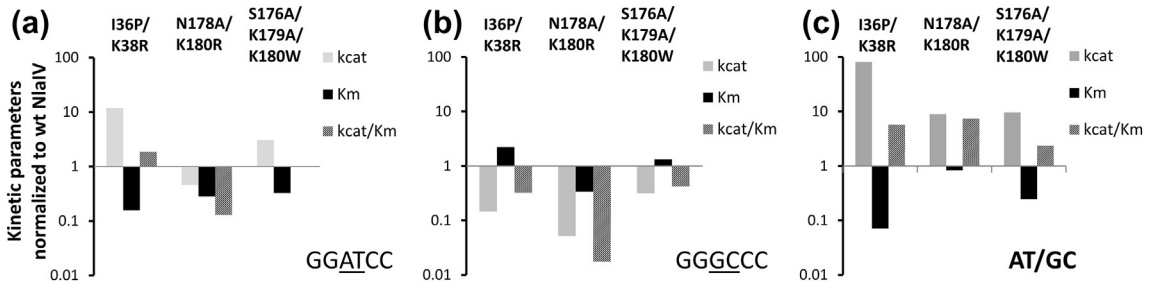

Fig. 3. Kinetic characterization of the most promising NlaIV variants. The values were normalized relatively to the parameters of the wild-type enzyme.

structure of NlaIV in complex with target DNA. We set up crystallization trials with NlaIV wild-type (except for the N-terminal GSH tag—the artifact of the thrombin catalyzed His_6 tag removal) and DNA oligoduplex containing the GGTACC target sequence. Crystals grew in the *P*6(2)22 space group and diffracted to 2.8 Å resolution on a synchrotron beamline. The structure was experimentally phased using a $Ta_6Br_{12}^{2+}$ derivative. Despite the presence of DNA in the crystallization mix, the electron density was present only for a single NlaIV protomer in the asymmetric unit but not for the DNA. The solvent content of the crystals equals almost 60%. The packing of the molecules is so loose that the presence of disordered DNA molecules cannot be excluded, but clearly, no substrate is specifically bound to the enzyme (Supplementary Fig. S5).

A comparison with other structures in the PDB performed with the DALI server [11] confirmed that NlaIV was structurally most similar to EcoRV (*Z*-score of 8.7). Other blunt-end cutting enzymes were only slightly more distant (*Z*-scores of 7.9 for HincII and SwaI). Overall, the crystal structure agreed with the previously published computational model of NlaIV [12]. However, shifts in the register in peripheral regions and substantial differences in the interface region were observed.

Dimerization region

The NlaIV protomers in the crystal are positioned close to the 2-fold axis. Application of the crystallographic symmetry completes the NlaIV dimer (Fig. 4). According to the PISA server [13,14], the dimerization interface extends over 1315 $Å^2$ and has a $\Delta^i G \sim -20$ kcal/mol (predicted solvation free energy gain upon dimer formation). According to the calculation, the probability of a chance occurrence of the contact is below 2%, confirming that the crystallographic interface is likely to be relevant also in solution.

The dimerization interface of the NlaIV shares some similarities with the one of EcoRV. It is composed of β-strands comprising residues 27–29 of the two protomers, that together with two other strands (residues 34–36) form a four-stranded anti-parallel β-sheet connected *via* main chain hydrogen bonds. The interface is completed by the downstream α-helices of the two protomers (residues 41–57). The analogous elements form the interface of EcoRV (four β-strands and two α-helices comprising residues 20–25, 30–32 and 37–58, respectively) (Supplementary Fig. S6). The additional structural fragment that contributes to the interface is located much further downstream in the sequence and is much more elaborate in EcoRV than in NlaIV (residues 135–137 in NlaIV and 143–156 in EcoRV).

The angle between the NlaIV protomers in the dimer present in the crystal is remarkably different from the one of the EcoRV–DNA complex [15]. The NlaIV dimer is wide-open, with active sites spaced too far apart for concerted cleavage of a DNA duplex. Therefore, we expect that the interface has to change upon DNA binding: presumably by a slight deformation of the β-sheet into a more barrel-like structure that would bring the active sites into closer proximity.

NlaIV active site

NlaIV has an active site typical for a PD-(D/E)XK restriction endonuclease, anchored in the context of the canonical α-β1-β2-β3 motif. As expected, there is a PD motif at the beginning of strand β1, where D73 plays the role of the "PD" aspartate. Downstream, there is the usual (D/E)XK motif, consisting in NlaIV of E87 and K89, the putative activator residue for the hydrolytic water molecule. The assignment of these residues is consistent with the earlier structure prediction. In PD-(D/E)XK restriction endonucleases, an acidic residue is expected in the α-helix immediately upstream of the α-β1-β2-β3 motif. This residue is E45 in the case of EcoRV, but somewhat

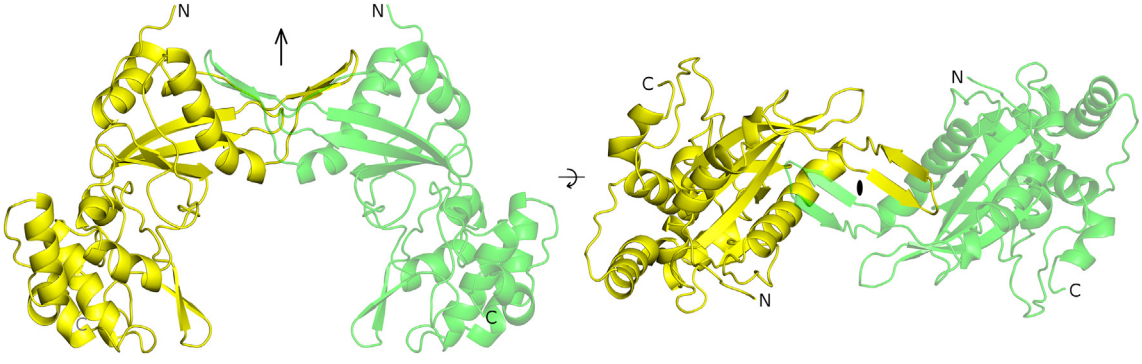

Fig. 4. Overall structure of the crystallographically observed NlaIV dimer. The protein dimer has been generated with the help of crystallographic 2-fold axis. The location of the axis is marked in the two orientations with an arrow and an oval.

atypically a glutamine, Q48, and not an acidic residue in NlaIV. The active-site residue in the α-helix typically coordinates water molecules that serve as ligands to one or two metal ions in the active site of a PD-(D/E)XK restriction endonucleases [15,16]. For this role, a glutamine is equally suitable as the more typically present glutamate. Alternatively, this residue might be unnecessary for the catalytic function of the NlaIV endonuclease, as observed for some other enzymes [17,18]. Since metal ions are not seen in the NlaIV active site, presumably because of the absence of the substrate DNA, this question remains open (Fig. 5).

Recognition of the central base pairs

Structure comparisons performed with the help of the DALI server [11] confirm the conclusion previously reached based on the amino acid sequence alone [6] that among the restriction endonucleases of known structure, NlaIV (GGN/NCC) is most similar to the blunt end cutter EcoRV (GAT/ATC). Thus, one may assume that DNA could adopt a very similar conformation in the NlaIV- and EcoRV-specific complexes. However, this conclusion cannot be fully correct. In EcoRV, the central dinucleotide step is not in direct contact with any amino acid residues, and instead, the 5′-TA-3′ specificity is largely enforced by the DNA conformation alone. Modeling shows that the replacement of the TA by an RY step that has identical 2′-deoxyribose positions and glycosidic bond angles leads to clashes of the 6-amino or 6-keto groups of the purine bases. Moreover, substantial clashes are also expected for the exocyclic methyl group of thymine in the 5′-AT-3′ sequence context. The structural studies of the complexes of EcoRV endonuclease with non-cognate DNA with the 5′-AT-3′ dinucleotide in the center indicate that the alleviation of the steric conflicts leads to the destabilization of the active site and a 10^4-fold difference in the cleavage efficiency [21]. In contrast, the substitution of the TA step with the CI (or 5mCI) step results in an almost unperturbed conformation of the EcoRV–DNA complex [22]. Thus, the EcoRV complex-derived DNA conformation in the case of NlaIV should not be compatible at least with the GGRYCC subset of the GGNNCC NlaIV target sequences (Supplementary Fig. S7).

Recognition of the outer base pairs

The recognition of the outer base pairs of the target sequence is not likely to be conserved between EcoRV and NlaIV and is thus difficult to predict. The only residue that is conservatively substituted according to the structure-based sequence alignment and might fulfill the same role is Q69 in NlaIV that corresponds to N70 in EcoRV. This residue donates a hydrogen bond to the O2 group of the thymine in the non-central A:T pair of the G**A**T/A**T**C target sequence of EcoRV. NlaIV recognizes a G:C pair in this position (G**G**N/N**C**C), and thus, the same hydrogen bond can also be formed. The other hydrogen bonds responsible for the recognition of the target base pairs are provided by the side chain and main chain atoms of the 182–188 fragment of EcoRV. This fragment is conserved neither in length nor in sequence in NlaIV, and thus, the specificity providing residues cannot be directly deduced from the structure-based sequence alignment. Based on the proximity to the target DNA, one can only suspect that the fragments of the enzyme that are responsible for base and backbone contacts with the DNA are located in three regions of the protein, namely, residues 68–69, 100–105 and 160–170. One of these regions (T68-Q69) was included in the mutagenesis, but did not yield variants with altered specificity, most likely due to the contacts with the other bases of the target sequence that needed to be preserved (Supplementary Figs. S8 and S9).

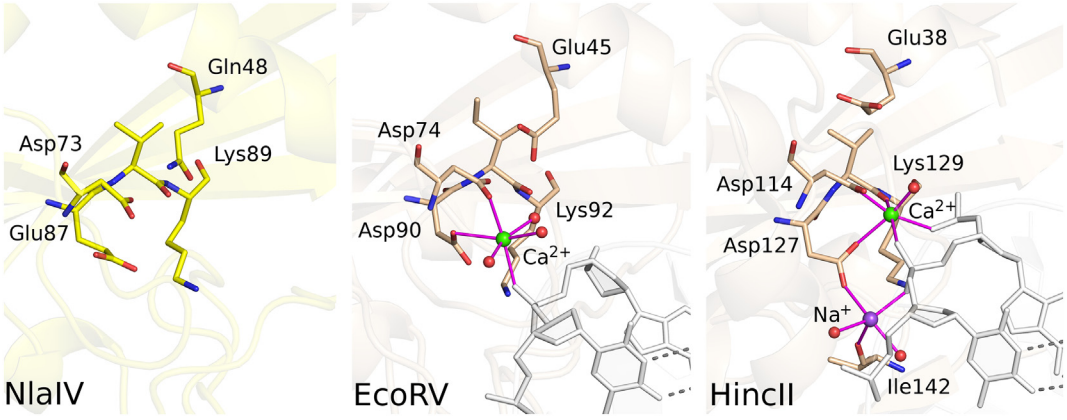

Fig. 5. Comparison of the NlaIV, EcoRV [19] and HincII [20] active sites.

Mapping of mutagenized regions onto the NlaIV–DNA complex model

The sites of random mutagenesis were selected based on a model of the NlaIV–DNA complex generated before the structure of the apo form of the enzyme was determined [6]. The model was obtained with the help of the EcoRV–DNA structure [15] that was the best template at the time. Surprisingly, the amino acid exchanges in regions that were correctly predicted to be located in the immediate vicinity of the substrate DNA did not result in the desired specificity change. The best results were instead obtained by substitutions that proved to be located in the parts of the protein that do not directly contact the substrate (at least for the DNA length present in the EcoRV crystal and thus also in the NlaIV model). The sites of substitutions that led to the most prominent specificity alterations are instead positioned either on the dimerization interface between the two protomers (residues 36 and 38) or in the region close to the exit tunnel of the DNA (residues 176–180). If the current model for the NlaIV complex with DNA is correct, the observed specificity change must be attributed to indirect effects of the introduced substitutions on the NlaIV–DNA interaction in the course of catalysis, which is reflected in the dominant role of the k_{cat} contribution (Fig. 6, and Supplementary Figs. S8 and S10).

Discussion

Building on natural specificity preferences

The recognition sequence of NlaIV has been described as GGN/NCC, but the naturally occurring enzyme nevertheless exhibits a 17-fold preference (judging from k_{cat}/K_M) for the AT step over the GC step in the center. The directed evolution approach to achieve GGA/TCC specificity could therefore build on pre-existing preference. Surprisingly, the evolution of GGA/TCC specificity also increased the activity of NlaIV toward GGT/ACC. This could indicate that anti-selection against enzyme variants cleaving a target sequence with the SS step at the center was more efficient than positive selection for enzymes with activity toward the GGA/TCC substrate. Formally, anti-selection against specificity for a 5′-SS-3′ step at the center should result in a preference for 5′-WW-3′. For NlaIV, this indeed appeared to occur and the selectivity for palindromic GGW/WCC target sequences was obtained.

Specificity changes are mostly k_{cat}, not K_M effects

It was initially surprising to us that the changes in specificity were primarily caused by differences in k_{cat} values. The result need not indicate that specificity alterations are more easily achieved by k_{cat} than by K_M changes, but could be the consequence of the specific selection conditions in our experiment. Reaction rates depend on the k_{cat}/K_M ratio in the regime where the substrate is much more abundant than the enzyme, and concentrations are low compared to K_M, so that changes in the latter control the concentration of productive complexes in solution. In the emulsion droplets, at least in theory there should only be a single DNA template (with two potential target sites), which can be used to generate multiple mRNA transcripts and many protein copies. Although we have not directly measured this, it is likely that the emulsion experiments are carried out in the regime of enzyme excess over substrate, and possibly also in conditions where the enzyme concentration may exceed the low (tight) K_M for the formation of specific restriction endonuclease DNA complexes. At least in the latter regime, reaction rates are only dependent on k_{cat}, and it therefore should

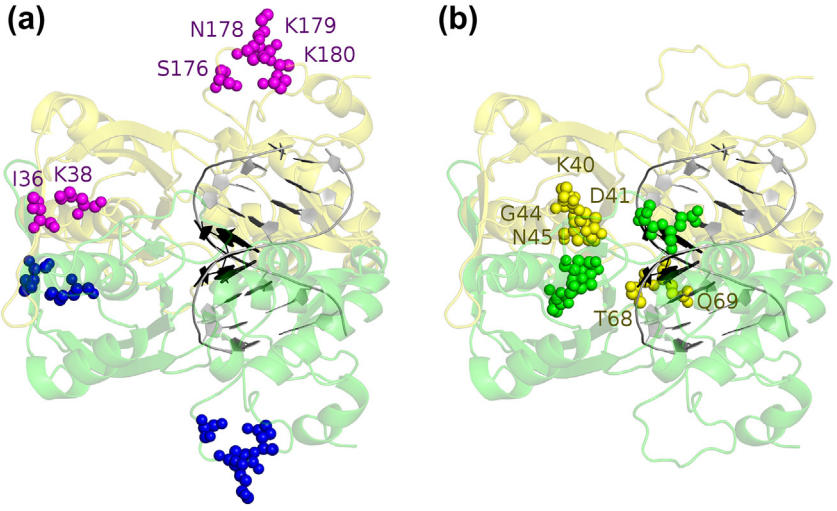

Fig. 6. Localization of the mutagenesis regions in the NlaIV–DNA model. The residues that have been randomly mutagenized are indicated as spheres: (a) the sites that when mutated led to the specificity change and (b) the sites that led to at most minor effects.

not come as a surprise that selection operates on the k_{cat} and not on K_M values.

Comparison of the structural basis of specificity for the central dinucleotide in blunt end cutters

In this work, we have added additional specificity to NlaIV, for the central two base pairs. Within the GGN/NCC context, the engineered preference is novel, but there are of course naturally occurring blunt-end cutting restriction endonucleases with specificity for the two base pairs. REBASE [23] lists altogether 224 such enzymes. Among these, 114 have a GC, 58 CG, 37 TA, 12 and AT step. There are also three examples (HinJCI, HincII and HindII) for the YR step, but no examples for a RY, WW, SS or any other degenerate dinucleotide steps.

The crystal structures of dimeric PD-(D/E)XK restriction enzymes in complex with substrate DNA are available for all but the AT step. For the TA step, the examples are EcoRV (GA**T/A**TC) [15] and SwaI (ATT**T/A**AAT) [12], both known for complex modes of recognition of the central step. For the CG step, NaeI (GC**C/G**GC) is so far the only example [24]. For the GC step, the examples include PvuII (CA**G/C**TG) [25] and ThaI (C**G/C**G) [16]. Both enzymes recognize the central step in a fairly canonical manner by a combination of shape selection and sequence-specific hydrogen bonding. The AT step is the only one without clearly documented endonuclease–DNA complex. DpnI (G**6mA/T**C) has a 6mAT step at the center of its target sequence, but the enzyme is monomeric and cleaves only methylated DNA [26].

There are four palindromic dinucleotide steps involving 2-fold degenerate base recognition, RY, YR, WW and SS. HincII [27] is an example enzyme that exhibits semi-degenerate specificity for the central step. The DNA in the HincII-DNA complex is strongly kinked, with major base pair roll and purine cross-stacking at the center. This DNA conformation excludes purines in the first position and thus selects for the YR step. The DNA conformations in the EcoRV [15] and NaeI [24] complexes are similar, and indeed, the central base pair steps of their target sequences are compatible with the YR consensus. However, both enzymes select among the structurally "allowed"

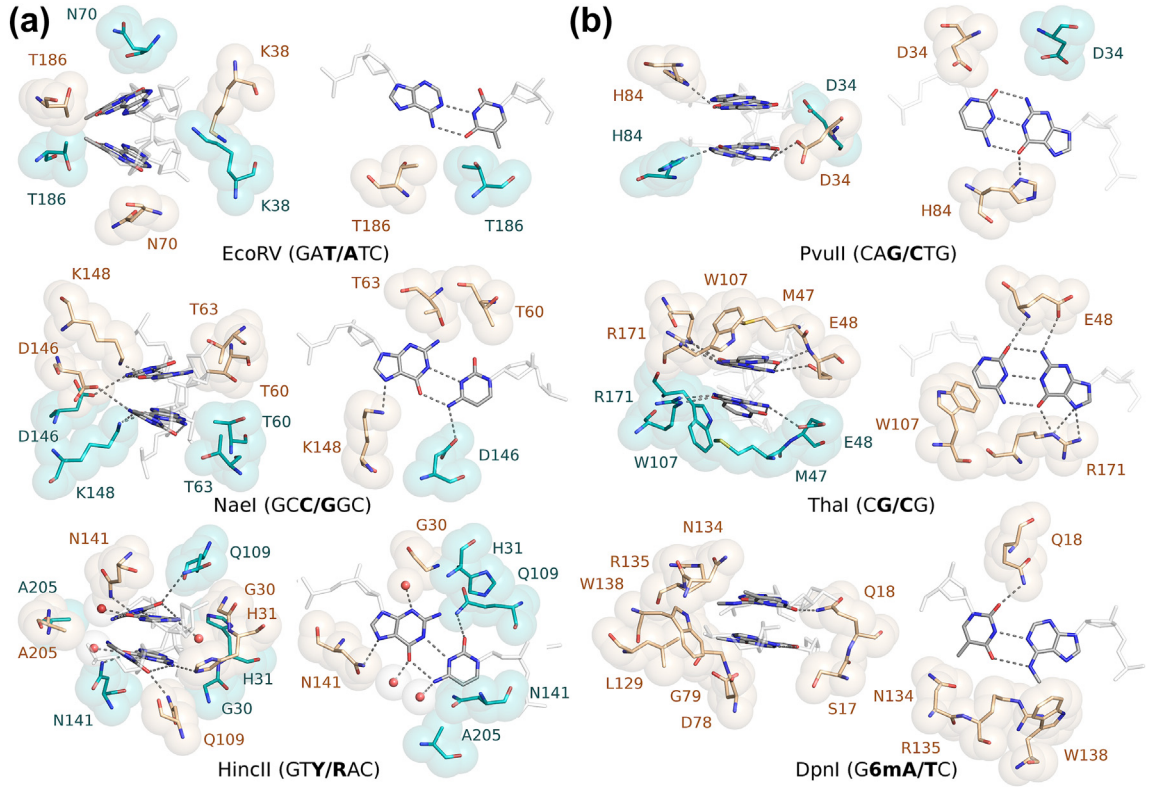

Fig. 7. Comparison of the central base pair recognition in the available blunt end cutting restriction endonuclease DNA complexes. Enzymes specific for (a) pyrimidine/purine and (b) purine/pyrimidine step in the center are presented in two orientations. The left panels indicate the residues surrounding both central base pairs, and the right panels indicate only one of them (for all enzymes but DpnI, the recognition of the target site is symmetric, and thus, the second base pair is bound analogously). A 4 Å distance cutoff was used to locate the residues surrounding the central base pairs.

possibilities. EcoRV achieves specificity for a TA step purely by indirect readout. NaeI also uses hydrogen bonds (involving D146 and K148 residues). PvuII [28] and ThaI [16] that bind kinked DNA without the roll at the center have targets with the central GC, which is not compatible with the YR consensus. Together, the data make a strong case that DNA structure can help to select a YR dinucleotide step (Fig. 7).

Selection for WW or SS steps has not yet been seen for blunt end-cutting enzymes. However, the structural basis for W or S recognition in the context of pseudo-palindromic sequences is very well understood. The distinction can be realized *via* probing of the base pairing strength by nucleotide flipping, as exemplified by the PspGI endonuclease [29,30]. Alternatively, the enzymes can distinguish W and S pairs by hydrogen bonds and/or shape selection. Inspection of hydrogen bonding requirements shows that they are base pair specific on the major groove side, and thus, this side is largely unsuitable for degenerate recognition [31]. In contrast in the center of the minor groove, the 2-amino group of guanine is almost in the same position for the G:C and C:G pairs and absent for A:T and T:A pairs. By testing for the presence or absence of the 2-amino group, enzymes can distinguish W from S, as observed for BcnI, MvaI or Hpy99I endonucleases [32–34].

The structural and biochemical data presented in this work suggest that the engineered NlaIV variants do not use probing of the central minor groove position or nucleotide flipping to achieve the semi-degenerate WW specificity. Mutations in the dimer interface region could affect steric crowding in the vicinity of the DNA in such a manner that 2-amino groups in the central minor groove position would be better rejected than for the wild-type NlaIV. However, since the specificity changes in NlaIV are primarily determined by the k_{cat} and not K_M value, such interpretation of the experimental data cannot be fully correct. Moreover, we do not see any residues coming closer than 4 Å to the central minor groove neither in the EcoRV–DNA structure [19] nor in our NlaIV–DNA complex model. Therefore, the preference of the engineered NlaIV variants likely results from the indirect influence of the DNA sequence on the active site conformation as implied before for the non-cognate target sequences of EcoRV [22] and in line with the dominant k_{cat} contribution.

Materials and Methods

Library construction

The NlaIV REase (NlaIV.R) expression cassette containing all necessary elements for *in vitro* transcription/translation/selection was constructed using pNlaRET28 plasmid (Fig. 1 and Supplementary Fig. S1) [6]. Three unique restriction sites (SalI, EcoRI and Eco52I) were introduced as silent mutations upstream of the intended mutagenesis regions. Moreover, the NlaIV site 77 bp after NlaIV.R stop codon was removed so that the whole sequence did not contain any NlaIV recognition sites (for primers, see Supplementary Table S2). Mutagenic primers were generated with ASM-800 DNA synthesizer (Biosset Ltd.) using split and mix strategy [35,36]. Each synthesis was initiated in all eight columns. At every split and mix step, the columns were opened, solid support resin was mixed in an Eppendorf tube and new columns were filled with mixed resin in the proportions dictated by the primer design (assuming 90% resin recovery at each mixing step). Synthesis products were purified on a Glen-Pak™ DNA Purification Cartridge (Glen Research), and their purity and integrity were assayed in denaturing polyacrylamide gel electrophoresis. Other primers were purchased from commercial suppliers (Sigma).

Libraries of mutagenized expression cassettes were generated in two-step PCR reactions (Supplementary Fig. S1). In the first step, two amplification reactions were performed. A non-mutagenic reaction amplified the cassette from the 5′ end up to a unique restriction site preceding mutagenic region. A second reaction using a 5′ mutagenic primer then amplified the expression cassette from the unique restriction site to the 3′ end of the cassette. Both reaction products were cleaved at the same site and ligated. The ligation product of the desired size (containing joined parts generated in the first step) was gel-purified and used as a template at the second amplification step. In this step, the GGSSCC and GGATCC flanking sequences were added for positive and negative selection. They were introduced using a non-biotinylated primer with the GGSSCC NlaIV recognition sequence and a 5′-biotinylated primer with the GGATCC sequence.

A variant with two altered regions combined (I36P/K38R/N178A/K180R) was constructed by replacing the Eco52I–XhoI segment of the I36P/K38R expression construct with the corresponding segment of the N178A/K180R coding sequence.

In vitro compartmentalization/transcription/translation and selection of specificity variants

Fifty microliters of the *in vitro* transcription/translation reaction mix (Rapid Translation System RTS 100, *Escherichia coli* HY Kit, Roche) containing 1.7 fmol of the expression cassette (1.18 ng) and supplemented with 5 mM $MgCl_2$ were emulsified in 0.95 ml of the oil/surfactants mix as described [37]. After 6 h of incubation at 30 °C, the samples were transferred to 37 °C for 3 h, the emulsion was broken by centrifugation and chloroform extraction, and DNA was precipitated from the aqueous phase. Precipitated DNA was dissolved and mixed with 10 µl of streptavidin

magnetic beads (New England Biolabs) and incubated 5 min at room temperature. After separation of the beads on a magnetic stand, the supernatant was collected and used for the re-amplification of selected clones. Two consecutive PCR reactions were used. In the first reaction, a primer selective for an intact 5′ end (with the uncleaved GGSSCC NlaIV sequence) was used in combination with the 3′ primer recreating the original 3′ end of the expression cassette (with the GGATCC NlaIV site). In the second PCR reaction, a primer randomizing the NlaIV site at the 5′ end to GGSSCC and a 3′ end primer with biotin on the 5′ end were used to recreate the expression cassette for use in another *in vitro* compartmentalization/selection round. This process was iterated five times, and aliquots of the selected variants (products of the first PCR reaction after streptavidin capture) were collected from each cycle (Supplementary Fig. S2).

Selected variants were cloned as NcoI–XhoI fragments into pET28a vector (Novagene) and transformed into the *E. coli* strain ER2566 (New England Biolabs) containing the pM.NlaIVAC plasmid expressing the NlaIV methyltransferase. Transformants were selected on LB with kanamycin, chloramphenicol and 1% glucose and used for expression of the NlaIV endonuclease. The expression was induced at mid-logarithmic phase of a culture with 1 mM IPTG, and cells were harvested 5 h later. Pellets from 15 ml expression cultures were used to purify His-tagged NlaIV endonuclease on 20 µl of HIS-Select Nickel Affinity resin (Sigma) in small-scale batch protocol according to the manufacturer's suggestions. Purified protein was eluted from the resin twice with 20 µl of the elution buffer. Two microliters of the protein sample (equivalent to 0.75 ml of the induced culture) was used in 2-h-long cleavage of 0.5 µg of lambda DNA. If the banding pattern generated by a variant was distinct from that generated by the wild-type enzyme, the variant was further characterized by a cleavage assay on a panel of 10 substrates (444-bp DNA fragments derived from pKSII Bluescript vector (Stratagene)) containing all combinations of the central dinucleotide in the NlaIV recognition site. For kinetic measurements, selected variants were purified from 2 l of induced cultures as described before [6].

Cleavage assay

All cleavage assays were performed at 37 °C in Fermentas Tango buffer [33 mM Tris-acetate (pH 7.9), 10 mM magnesium acetate, 66 mM potassium acetate and 0.1 mg/ml BSA]. Real-time kinetics of DNA cleavage were followed for 5 min in Shimidzu spectrofluorometer [38] using a hairpin substrate with a single NlaIV site close to the base of the stalk (Supplementary Fig. S3). The oligonucleotide substrate contained a ROX label at the 5′ end and a BHQ2 quencher at the 3′ end; therefore, ROX fluorescence in the annealed molecule was quenched. Upon cleavage, the 7-bp fragment containing the labeled ends of the substrate denatured at the reaction temperature (37 °C, calculated Tm = 6 °C) resulting in the separation of the fluorophore and quencher and a fluorescence increase. To convert fluorescence units into the molar concentration of the product, the increase of fluorescence for 200 nM substrate upon overnight cleavage with wt enzyme was recorded. Initial cleavage rates at varying substrate concentrations (5–200 nM) were measured and used to calculate Michaelis–Menten kinetic constants. Measurements were made in triplicate.

Protein expression and purification for crystallization

The NlaIV variant with the N-terminal MGSSHHHHH HSSGLVPRGSH was used for crystallographic purposes. The protein was expressed from pNlaNHis plasmid that was constructed by cloning of the NdeI–XhoI fragment of the amplification product of the NlaIV ORF from pNlaRET plasmid with NlaNHisF and NlaNHisR primers into the pET28a vector digested with the same enzymes. The protein was expressed as described above and purified on a Ni-NTA column equilibrated with the buffer containing 15% glycerol, 50 mM Hepes (pH 7.5), 200 mM NaCl, 10 mM imidazole, 10 mM β-mercaptoethanol and 0.1% Triton X-100. The column was washed twice: with the above buffer but with the amount of salt raised to 2 M NaCl, and then with the above buffer containing 20 mM imidazole and lacking Triton X-100. The elution was carried out with the buffer containing 15% glycerol, 50 mM Hepes (pH 7.5), 300 mM NaCl, 250 mM imidazole and 10 mM β-mercaptoethanol. Pooled peak fractions were dialyzed against buffer containing 15% glycerol, 50 mM Hepes (pH 7.5), 10 mM β-mercaptoethanol and 50 mM KCl, and the histidine tag was removed with thrombin according to the standard protocol leaving the MIK sequence at the N-terminus. The obtained protein was further purified on an SP column in the 50 mM–1 M KCl gradient in the same buffer. The buffer of the purified protein preparation was exchanged on PD-10 column to 15% glycerol, 50 mM Hepes (pH 7.5), 50 mM KCl, 10 mM DTT and 5 mM $CaCl_2$.

Crystallization and structure determination

The NlaIV crystals were grown by sitting drop vapor diffusion method. The protein solution contained 10 mg/ml of the enzyme in 15% glycerol, 50 mM Hepes (pH 7.5), 50 mM KCl and 10 mM DTT and 5 mM $CaCl_2$. It was mixed with a double-stranded DNA oligonucleotide in 1:1 ratio (NlaIV dimer and dsDNA). The oligo was composed of the 5′-ATGGTACCTGC-3′ and 5′-CAGGTACCATG-3′ strands, which upon annealing produced single-nucleotide 3′ overhangs. Crystallization was performed by mixing of the

protein–DNA solutions in 1:3 ratio with the precipitant solution containing 2 M NaCl and 100 mM citric acid (pH 5). The crystals grew in the *P*6(2)22 space group. For cryo-protection, the crystals were soaked in the buffer containing 4 µl of 80% glycerol and 6 µl of the crystallization buffer. The native data set of 2.8 Å resolution was collected at 1.05 Å wavelength at the BW6 beamline of the DESY synchrotron (Hamburg, Germany). Since the structure solution with the help of molecular replacement proved difficult, the crystals were soaked with $Ta_6Br_{12}^{2+}$ solution [39]. A tiny amount of solid $Ta_6Br_{12}^{2+}$ was dissolved in the reservoir buffer. The soaking was performed for 1 min at room temperature. A two-wavelength MAD data set was collected at the MX 14.2 beamline of BESSY synchrotron (Berlin, Germany) at the 1.25489 Å (peak) and 1.25531 Å (inflection point) wavelengths (data to 3.0 and 3.2 Å resolution, respectively). The anomalous signal of the data collected at 1.25489 Å dropped below CC_{ano} of 25% at approximately 4.5 Å resolution. The data were processed with autoPROC [40] and STARANISO/XDS [41,42] programs. The structure was solved with the autoSHARP program [43]. Two-wavelength MAD was used to locate one strong heavy atom site with the help of the SHELXD program [44] [CC(all/weak) 42.10/29.30, PATFOM 21.63, CFOM 71.40]. The initial heavy atom parameters were submitted to heavy atom refinement and phasing in SHARP [45] [leading to overall FOM (acentric/centric) of 0.35/0.29], followed by density modification with SOLOMON [46]. A spherical averaged version of the $Ta_6Br_{12}^{2+}$ cluster was used to describe the heavy atom binding sites *via* special form factors as implemented in SHARP [45]. The resulting phases were further improved with the help of DM [47] and submitted to iterative density modification with PARROT [47] and model building with BUCCANEER [48] programs. The preliminary model was refined with BUSTER [49] and subsequently rebuilt with the ARP/wARP program [50]. The best parts of both models were manually combined and once again automatically rebuilt with ARP/wARP and BUCCANEER. The resulting structure was refined with the COOT [51] and REFMAC [52] programs, using isotropically truncated diffraction data. Data collection and refinement statistics are shown in Table 3. The final model coordinates and the corresponding structure factors were submitted to PDB with the 6QM2 accession code.

Table 3. Data collection and refinement statistics

Data collection statistics	
Space group	*P*6(2)22
Cell dimensions	
a (Å)	110.4
c (Å)	209.0
Wavelength (Å)	1.05
Resolution range (Å)	30–2.8
Lowest shell	30–8.1
Highest shell	2.97–2.8
Total reflections	142,779
Unique reflections	18,997
Completeness (%)[a]	98.5 (88.0, 98.6)
Multiplicity[a]	7.5 (6.5, 7.7)
Mean $I/\sigma I$[a]	28.9 (74.3, 1.7)
$CC_{1/2}$ (%)[a]	100 (100, 77.5)
R_{merge} (%)[a]	4.0 (1.9, 98.6)
R_{meas} (%)[a]	4.3 (2.1, 105.2)
Solvent content (%)	58
B_{iso} from Wilson (Å^2)	98.5
Refinement statistics	
Protein atoms excluding H	2086
Solvent molecules	125
R_{cryst} (%)	19.10
R_{free} (%)	22.10
RMSD bond lengths (Å)	0.007
RMSD angles (°)	1.083
Ramachandran favored region (%)	97.5
Ramachandran allowed region (%)	100
Molprobity clashscore	0

[a] Statistics for the highest- and lowest-resolution shell are indicated in brackets.

Molecular modeling

A preliminary model of NlaIV–DNA complex in a closed conformation (protein wrapped around the DNA, with active sites positioned for a double-stranded DNA cleavage) has been obtained computationally. Briefly, we generated a crude approximation of the EcoRV-like binding mode by superimposing NlaIV protomers onto the EcoRV–DNA dimer structure (PDB: 1EO3) [53]. The catalytic cores of individual NlaIV protomers and the dimerization region involving fragments of both protomers were moved as independent rigid bodies to maximize the fit of active sites, preserve the dimerization mode and maintain the continuity of the polypeptide chains. Subsequently, the crystallographically determined structure of NlaIV in an open conformation without the DNA was subjected to the computational normal mode analysis using the NOMAD server [54]. The procedure was applied to identify potential conformational changes involving all atoms of NlaIV dimer that could bring the molecule from the experimentally observed unbound conformation to the predicted bound conformation. The final model of NlaIV complex was constructed by splicing the catalytic domains and the dimerization region of NlaIV. DNA coordinates were copied from the EcoRV–DNA complex followed by computational mutagenesis to represent the NlaIV target sequence. The NlaIV model in the closed conformation was briefly optimized to improve local geometry and to reduce steric clashes. Extensive simulations of interactions were not attempted, as the model is not supposed to be accurate at the atomic level and is currently only used to illustrate the key assumptions about the mutual disposition of key structural elements and functionally important residues.

Accession numbers

The final model coordinates and the corresponding structure factors were submitted to Protein Data Bank with the accession number PDB ID: 6QM2.

Acknowledgments

We thank Drs. Uwe Mueller, Manfred Weiss and Gleb Bourenkov for beamtime allocation and assistance with crystallographic data collection at the BW6 beamline of the DORIS ring (MPG/DESY, Hamburg, Germany) and MX14.2 beamline of the BESSY storage ring (Berlin, Germany). We thank the tutors of the DLS-CCP4 Data Collection and Structure Solution Workshop and in particular Dr. Clemens Vonrhein (Global Phasing Ltd) for the help with structure determination. This work was supported by the grants from the Ministry of Science and Higher Education (0295/B/PO1/2008/34 to M.B., N N301 425038 to H.C., N301 100 31/3043 to K.S., DWM/N2/DFG/2008 to J.M.B.), from the Polish National Science Centre (NCN) (UMO-2011/02/A/NZ1/00052, UMO-2014/13/B/NZ1/03991 and UMO-2014/14/M/NZ5/00558 to M.B.) and by short-term EMBO fellowship to KS (ATSF 277.00-05) and from the Polish National Agency For Academic Exchange (PPI/APM/2018/1/00034).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2019.04.010.

Received 4 February 2019;
Received in revised form 14 March 2019;
Accepted 7 April 2019
Available online 14 April 2019

Keywords:
restriction endonuclease;
directed evolution;
in vitro compartmentalization;
specificity engineering;
crystal structure

Current address: W. Siwek, Instituto Gulbenkian de Ciencia, Rua da Quinta Grande, 6, 2780-156 Oeiras, Portugal.

References

[1] J.C. Samuelson, S.Y. Xu, Directed evolution of restriction endonuclease BstYI to achieve increased substrate specificity, J. Mol. Biol. 319 (2002) 673–683.

[2] T. Lanio, A. Jeltsch, A. Pingoud, On the possibilities and limitations of rational protein design to expand the specificity of restriction enzymes: a case study employing EcoRV as the target, Protein Eng. 13 (2000) 275–281.

[3] S. Schottler, C. Wenz, T. Lanio, A. Jeltsch, A. Pingoud, Protein engineering of the restriction endonuclease EcoRV—structure-guided design of enzyme variants that recognize the base pairs flanking the recognition site, Eur. J. Biochem. 258 (1998) 184–191.

[4] K. Skowronek, M.J. Boniecki, B. Kluge, J.M. Bujnicki, Rational engineering of sequence specificity in R.MwoI restriction endonuclease, Nucleic Acids Res. 40 (2012) 8579–8592.

[5] T. Lanio, A. Jeltsch, A. Pingoud, Towards the design of rare cutting restriction endonucleases: using directed evolution to generate variants of EcoRV differing in their substrate specificity by two orders of magnitude, J. Mol. Biol. 283 (1998) 59–69.

[6] A.A. Chmiel, M. Radlinska, S.D. Pawlak, D. Krowarsch, J.M. Bujnicki, K.J. Skowronek, A theoretical model of restriction endonuclease NlaIV in complex with DNA, predicted by fold recognition and validated by site-directed mutagenesis and circular dichroism spectroscopy, Protein Eng. Des. Sel. 18 (2005) 181–189.

[7] B.Q. Qiang, I. Schildkraut, 2 unique restriction endonucleases from Neisseria-Lactamica, Nucleic Acids Res. 14 (1986) 1991–1999.

[8] O.J. Miller, K. Bernath, J.J. Agresti, G. Amitai, B.T. Kelly, E. Mastrobattista, et al., Directed evolution by in vitro compartmentalization, Nat. Methods 3 (2006) 561–570.

[9] A.D. Griffiths, D.S. Tawfik, Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization, EMBO J. 22 (2003) 24–35.

[10] H.M. Cohen, D.S. Tawfik, A.D. Griffiths, Altering the sequence specificity of HaeIII methyltransferase by directed evolution using in vitro compartmentalization, Protein Eng. Des. Sel. 17 (2004) 3–11.

[11] L. Holm, L.M. Laakso, Dali server update, Nucleic Acids Res. 44 (2016) W351–W355.

[12] B.W. Shen, D.F. Heiter, K.D. Lunnen, G.G. Wilson, B.L. Stoddard, DNA recognition by the SwaI restriction endonuclease involves unusual distortion of an 8 base pair A:T-rich target, Nucleic Acids Res. 45 (2017) 1516–1528.

[13] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state, J. Mol. Biol. 372 (2007) 774–797.

[14] E. Krissinel, Crystal contacts as nature's docking solutions, J. Comput. Chem. 31 (2010) 133–143.

[15] F.K. Winkler, D.W. Banner, C. Oefner, D. Tsernoglou, R.S. Brown, S.P. Heathman, et al., The crystal-structure of Ecorv endonuclease and of its complexes with cognate and non-cognate DNA fragments, EMBO J. 12 (1993) 1781–1795.

[16] M. Firczuk, M. Wojciechowski, H. Czapinska, M. Bochtler, DNA intercalation without flipping in the specific ThaI–DNA complex, Nucleic Acids Res. 39 (2011) 744–754.

[17] H.G. Nastri, P.D. Evans, I.H. Walker, P.D. Riggs, Catalytic and DNA binding properties of PvuII restriction endonuclease mutants, J. Biol. Chem. 272 (1997) 25761–25767.

[18] G. Grabowski, G. Maass, J. Alves, Asp-59 is not important for the catalytic activity of the restriction endonuclease EcoRI, FEBS Lett. 381 (1996) 106–110.

[19] M.P. Thomas, R.L. Brady, S.E. Halford, R.B. Sessions, G.S. Baldwin, Structural analysis of a mutational hot-spot in the EcoRV restriction endonuclease: a catalytic role for a main

chain carbonyl group, Nucleic Acids Res. 27 (1999) 3438–3445.
[20] H.K. Joshi, C. Etzkorn, L. Chatwell, J. Bitinaite, N.C. Horton, Alteration of sequence specificity of the type II restriction endonuclease HincII through an indirect readout mechanism, J. Biol. Chem. 281 (2006) 23852–23869.
[21] D.A. Hiller, A.M. Rodriguez, J.J. Perona, Non-cognate enzyme–DNA complex: structural and kinetic analysis of EcoRV endonuclease bound to the EcoRI recognition site GAATTC, J. Mol. Biol. 354 (2005) 121–136.
[22] A.M. Martin, M.D. Sam, N.O. Reich, J.J. Perona, Structural and energetic origins of indirect readout in site-specific DNA cleavage by a restriction endonuclease, Nat. Struct. Biol. 6 (1999) 269–277.
[23] R.J. Roberts, T. Vincze, J. Posfai, D. Macelis, REBASE—a database for DNA restriction and modification: enzymes, genes and genomes, Nucleic Acids Res. 43 (2015) D298-D9.
[24] Q. Huai, J.D. Colandene, M.D. Topal, H.M. Ke, Structure of NaeI–DNA complex reveals dual-mode DNA recognition and complete dimer rearrangement, Nat. Struct. Biol. 8 (2001) 665–669.
[25] K. Balendiran, J. Bonventre, R. Knott, W. Jack, J. Benner, I. Schildkraut, et al., Expression, purification, and crystallization of restriction-endonuclease Pvuii with DNA containing its recognition site, Proteins Struct. Funct. Genet. 19 (1994) 77–79.
[26] K. Mierzejewska, W. Siwek, H. Czapinska, M. Kaus-Drobek, M. Radlinska, K. Skowronek, et al., Structural basis of the methylation specificity of R.DpnI, Nucleic Acids Res. 42 (2014) 8745–8754.
[27] N.C. Horton, L.F. Dorner, J.J. Perona, Sequence selectivity and degeneracy of a restriction endonuclease mediated by DNA intercalation, Nat. Struct. Biol. 9 (2002) 42–47.
[28] X.D. Cheng, K. Balendiran, I. Schildkraut, J.E. Anderson, Structure of Pvuii endonuclease with cognate DNA, EMBO J. 13 (1994) 3927–3935.
[29] R.H. Szczepanowski, M.A. Carpenter, H. Czapinska, M. Zaremba, G. Tamulaitis, V. Siksnys, et al., Central base pair flipping and discrimination by PspGI, Nucleic Acids Res. 36 (2008) 6109–6117.
[30] G. Tamulaitis, M. Zaremba, R.H. Szczepanowski, M. Bochtler, V. Siksnys, How PspGI, catalytic domain of EcoRII and Ecl18kI acquire specificities for different DNA targets, Nucleic Acids Res. 36 (2008) 6101–6108.
[31] N.C. Seeman, J.M. Rosenberg, A. Rich, Sequence-specific recognition of double helical nucleic acids by proteins, Proc. Natl. Acad. Sci. U. S. A. 73 (1976) 804–808.
[32] M. Sokolowska, M. Kaus-Drobek, H. Czapinska, G. Tamulaitis, R.H. Szczepanowski, C. Urbanke, et al., Monomeric restriction endonuclease BcnI in the apo form and in an asymmetric complex with target DNA, J. Mol. Biol. 369 (2007) 722–734.
[33] M. Kaus-Drobek, H. Czapinska, M. Sokolowska, G. Tamulaitis, R.H. Szczepanowski, C. Urbanke, et al., Restriction endonuclease MvaI is a monomer that recognizes its target sequence asymmetrically, Nucleic Acids Res. 35 (2007) 2035–2046.
[34] M. Sokolowska, H. Czapinska, M. Bochtler, Crystal structure of the beta beta alpha-Me type II restriction endonuclease Hpy99I with target DNA, Nucleic Acids Res. 37 (2009) 3799–3810.
[35] S.M. Glaser, D.E. Yelton, W.D. Huse, Antibody engineering by codon-based mutagenesis in a filamentous phage vector system, J. Immunol. 149 (1992) 3903–3913.
[36] S.J. Lahr, A. Broadwater, C.W. Carter Jr., M.L. Collier, L. Hensley, J.C. Waldner, et al., Patterned library analysis: a method for the quantitative assessment of hypotheses concerning the determinants of protein structure, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 14860–14865.
[37] R. Williams, S.G. Peisajovich, O.J. Miller, S. Magdassi, D.S. Tawfik, A.D. Griffiths, Amplification of complex gene libraries by emulsion PCR, Nat. Methods 3 (2006) 545–550.
[38] S.S. Ghosh, P.S. Eis, K. Blumeyer, K. Fearon, D.P. Millar, Real time kinetics of restriction endonuclease cleavage monitored by fluorescence resonance energy transfer, Nucleic Acids Res. 22 (1994) 3155–3159.
[39] J. Knablein, T. Neuefeind, F. Schneider, A. Bergner, A. Messerschmidt, J. Lowe, et al., Ta6Br(2+)12, a tool for phase determination of large biological assemblies by x-ray crystallography, J. Mol. Biol. 270 (1997) 1–7.
[40] C. Vonrhein, C. Flensburg, P. Keller, A. Sharff, O. Smart, W. Paciorek, et al., Data processing and analysis with the autoPROC toolbox, Acta Crystallogr. D Biol. Crystallogr. 67 (2011) 293–302.
[41] I.J. Tickle, C. Flensburg, P. Keller, W. Paciorek, A. Sharff, C. Vonrhein, et al., STARANISO, Global Phasing Limited, 2018.
[42] W. Kabsch, Xds, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 125–132.
[43] C. Vonrhein, E. Blanc, P. Roversi, G. Bricogne, Automated structure solution with autoSHARP, Methods Mol. Biol. 364 (2007) 215–230.
[44] T.R. Schneider, G.M. Sheldrick, Substructure solution with SHELXD, Acta Crystallogr. D Biol. Crystallogr. 58 (2002) 1772–1779.
[45] G. Bricogne, C. Vonrhein, C. Flensburg, M. Schiltz, W. Paciorek, Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0, Acta Crystallogr. D Biol. Crystallogr. 59 (2003) 2023–2030.
[46] J.P. Abrahams, A.G. Leslie, Methods used in the structure determination of bovine mitochondrial F1 ATPase, Acta Crystallogr. D Biol. Crystallogr. 52 (1996) 30–42.
[47] K. Cowtan, Recent developments in classical density modification, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 470–478.
[48] K. Cowtan, The Buccaneer software for automated model building. 1. Tracing protein chains, Acta Crystallogr. D Biol. Crystallogr. 62 (2006) 1002–1011.
[49] E. Blanc, P. Roversi, C. Vonrhein, C. Flensburg, S.M. Lea, G. Bricogne, Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT, Acta Crystallogr. D Biol. Crystallogr. 60 (2004) 2210–2221.
[50] G. Langer, S.X. Cohen, V.S. Lamzin, A. Perrakis, Automated macromolecular model building for x-ray crystallography using ARP/wARP version 7, Nat. Protoc. 3 (2008) 1171–1179.
[51] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 486–501.
[52] G.N. Murshudov, A.A. Vagin, E.J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method, Acta Crystallogr. D Biol. Crystallogr. 53 (1997) 240–255.
[53] N.C. Horton, B.A. Connolly, J.J. Perona, Inhibition of EcoRV endonuclease by deoxyribo-3′-*S*-phosphorothiolates: a high-resolution x-ray crystallographic study, J. Am. Chem. Soc. 122 (2000) 3314–3324.
[54] E. Lindahl, C. Azuara, P. Koehl, M. Delarue, NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis, Nucleic Acids Res. 34 (2006) W52–W56.